



AN APPROACH OF SECURITY IN E-COMMERCE WITH WEB MINING FRAMEWORK

Afrin M Sheikh¹ | Sumitra Menaria²

¹ Master of Technology, Parul Institute of Engg. And Technology, Parul University, Vadodara.

² Assistant Professor, Parul Institute of Engg. And Technology, Parul University, Vadodara.

ABSTRACT

Web Mining is an extensible and very important area of the Data Mining which deals with the extraction of interesting knowledge from the World Wide Web, It can be categorised majorly into three types i.e. web content mining, web structure mining and web usages mining. This paper is based on e-commercial web sites how to use web mining techniques for providing security and better access on e-commerce web sites. The connection between web mining, security and e-commerce analyzed based on user behavior on web. Based on customer behavior different kind of web mining algorithms like page rank algorithm and trust rank algorithm is used for developing web mining framework in e-commerce web sites. We have developed Encryption technique to provide security on e-commerce web site.

General Terms: Web mining techniques, E-commerce.

KEYWORDS: E-commerce, Security, data mining, web mining.

1. INTRODUCTION

The information on the internet is in the form of static and dynamic web pages of various areas from education, industry to every walk of life including blogs. As per the web sites' survey more than 160,000,000 web sites are having inter, intra linked web pages. The speed of web information is rapid. The way the web sites and web pages are accessed, it is useful from the business perspective for giving future directions for decision making. Data mining (sometimes called data or knowledge discovery) is the process of analysing data from different perspectives and summarizing it into useful information.

The Web Mining is an application of the data mining techniques to find interesting and potentially useful knowledge from web data. Users face many problems due to the huge volume of information that is consistently growing. In particular, Web users have issues in getting the correct information due to low precision. For example, if a user wants to get any information by using Google and other search engines, it will provide not only Web contents dealing with this topic, but a series of irrelevant information, so called noise pages, resulting in difficulties for users in obtaining necessary information.

2. WEB MINING FRAMEWORK SYSTEM

Web mining makes the use of data mining techniques to automatically discover and extract knowledge from the web documents. web mining provides the information service centre in various fields like news, e-commerce, and advertisement, government, education, financial management, education, etc. We have developed Web mining framework for evaluating e-commerce web sites.

In general, Web mining tasks can be classified into three categories:

1. Web content mining,
2. Web structure mining and
3. Web usage mining[3].

Web Content Mining

Today, there are several billions of HTML documents, pictures, images and other multimedia files available via the internet and the number is still rising. Retrieving the interesting and necessary contents has become a very difficult and important task taking into consideration the impressive variety of the Web. Web content consists of several types of data such as text data, images, audio or video, Unstructured/structured records such as lists or tables and hyperlinks. Web content mining can be defined as the scanning and mining of text, graphs and pictures from a Web page to find out the significance of the content to the search query.

It deals with finding useful information or knowledge from the web page contents. Margaret H. Dunham [11] stated Web Content Mining can be thought of as extending the work performed by basic search engines. The primary Web resources that are mined in Web content mining are individual pages. Information Retrieval is one of the research areas that provide a range of popular and effective, mostly statistical methods for Web content mining.

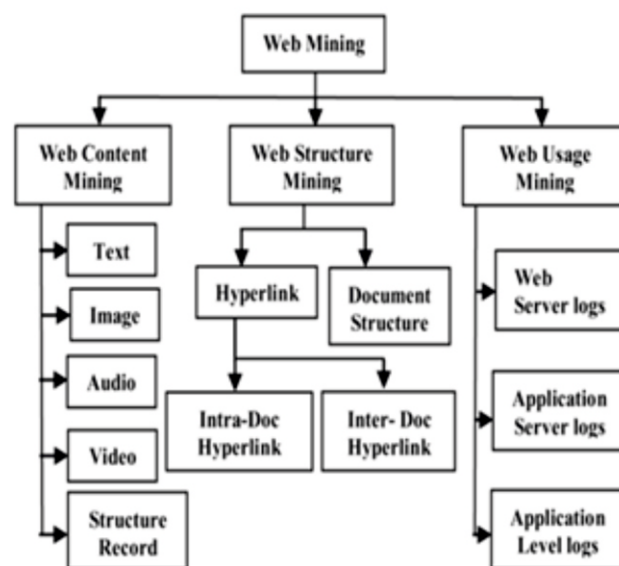


Fig.1 Taxonomy of Web mining

Web Structure Mining (Web Linkage Mining)

It deals with discovering and modelling the link structure of the web. Web information retrieval tools make use of only the text available on web pages but ignoring valuable information contained in web links. Its aims to generate structural summary about web sites and web pages. The main focus of web structure mining is on link information.

As the size and complexity of the websites expands dramatically, it has become more and more challenging aspect to design websites on which the web surfers can easily find the information they seek. Fang and Sheng [14] address the design of the portal page of a web site. They try to maximize the efficiency, effectiveness, and usage of a web site's portal page by selecting a limited number of hyperlinks from a large set for the inclusion in a portal page. Based on relationships among the hyperlinks (i.e. structural relationships that can be extracted from a web site and access relationship that can be discovered from a web log), they proposed a heuristic approach to hyperlink selection called Link Selector. Instead of clustering user navigation patterns by means of a Euclidean distance measure, Hay et al.[15] use the Sequence Alignment Method (SAM) to partition users into clusters, according to the order in which web pages are requested and the different lengths of clustering sequences. Web Structure Mining plays a vital role with various benefits including quick response to the web users, reducing lot of HTTP transactions between users and server.

The structure of a typical Web graph consists of Web pages as nodes , and hyperlinks as edges connecting related pages. Web Structure Mining is the process of discovering structure information from the Web. This can be further divided into two kinds based on the kind of structure information used.

Hyperlinks: A Hyperlink is a structural unit that connects a location in a Web page to different location, either within the same Web page or on a different Web page. A hyperlink that connects to a different part of the same page is called an Intra-Document Hyperlink, and a hyperlink that connects two different pages is called an Inter-Document Hyperlink.

Document Structure: In addition, the content within a Web page can also be organized in a tree-structured format, based on the various HTML and XML tags within the page. Mining efforts here have focused on automatically extracting document object model (DOM) structures out of documents

Web Usage Mining

It deals with understanding the user behaviour how they interacting with the web or with a website. One of the aims is to obtain information that may assist web site reorganization or assist the site adaptation to better suit the user. Web usage mining model is a kind of mining to server logs and web logs and its aim is getting useful users' access information in logs to make sites can perfect themselves with appropriate users' requirements, serve users better and get more economy benefits.

Web usage mining analyzes the information about visited Web pages that are saved in log files of Internet servers in order to finding interesting patterns previously unknown and potentially useful. It can be described as the mining applying data mining techniques on Web access logs to optimize web site for users. There are many web log analysis tools available to mine data from log record on web page. Log record contains plenty of useful information such as URL, IP address and time and so on. Analyzing and discovering Log could help organizations to find more potential customers, pages popularity (number of times a page has been visited) etc that can help in reorganizing the web site for fast and easy customer access.

3. SECURITY PERSPECT

Lack of trust is one of the main reasons which can make e-commerce less attractive because of the fear of credit card number or sensitive information being stolen[12]. The increasing number of the web security attacks causes fears to consumers that resulted in lack of trust. Hence, many businesses and internet users are reluctant to use the new technology.

According to the largest internet security company McAfee , almost half of consumers had terminated an order or due to security fears. Even in an attempt to get a good deal, 63% consumers will refuse to purchase from a Website that does not show a Trustmark or security policy[7]. Usually, e-commerce firms seek to get trust of their users by creating and advertising new security strategies, but the security threat is still growing and affecting e-commerce firms negatively.

I. E-Security Issues and Trust

Threats can be made either through network and data transaction attacks, or via unauthorized access by means of defective authentication. For customers, it must be recognized that economic hardship encompasses damages to privacy as well as theft, of credit information and authentication issues for consumers will be overturned; as in whether the Web site is „real“ rather than whether the purchaser's identity is real. This modified definition explains the security threats from a consumer's point of view. Security in B2C electronic commerce is reflected in the technologies used to secure customer data. Security concerns of consumers may be addressed by many of the same technology protections as those of businesses, such as encryption and authentication[1].

Because of all these security issues there is a great need of web security. Therefore the proposed system will implement the security by implementing the encryption ciphering techniques.

4. OUR CONTRIBUTION

This paper proposes a system model which is very useful in e-commerce applications and its security. This system involves integration of web mining framework with an e-commerce application involving the classification technique and also with Encryption technique for providing security. This integration facilitate e-store owner to improve the features and security. There are many areas where data mining can be very helpful when integrating with e-commerce.

Classification is one of the Data Mining techniques that is mainly used to analyze a given data set and takes each instance of it and assigns this instance to a particular class such that classification error will be least. It is used to extract models that accurately define important data classes within the given data set.

Classification is a two step process. During first step the model is created by applying classification algorithm on training data set then in second step the extracted model is tested against a predefined test data set to measure the model trained performance and accuracy. So classification is the process to assign class label from data set whose class label is unknown.

5. CONCLUSIONS

In the proposed model web mining integrated with the electronic commerce application to improve the performance of e-commerce applications and provided with better security. First we have discussed some important taxonomy of web mining which are used in data mining. After that we explained the proposed architecture which contains mainly user identification phase which involves user authentication task by applying encryption techniques and makes use of classification technique .

6. REFERENCES

- [1]. R.Manjusha, Research Scholar, Dr. R.Ramachandran, “Web Mining Framework for Security in E-commerce”, IEEE - International Conference on Recent Trends in Information Technology, ICRTIT 2011
- [2]. Wang Yong gui, Jia Zhen, “Research on Semantic Web Mining” ,2010 International Conference On Computer Design And Applications (ICDDA) The Internet and Management, Vol. 10, No.3, 2002, p 1 - 14
- [3]. Jinhua sun, Yanqi Xie, “A Web Data Mining Framework for E-commerce Recommender Systems”, 978-1-4244-4507-3
- [4]. SaiMing Au, "A study of application of web mining for e-commerce: tools and methodology", International Journal of the Computer, the Internet and Management 10, no. 3 (2002): 1-14.
- [5]. “Web Mining In E-Commerce”, Istrate Mihai University of Pitesti Faculty of Mathematics and Informatics No.1, Stadionului, Ticleni, Gorj, 215600, Romaniamihai@france@yahoo.com +40 745 775 935
- [6]. <https://arxiv.org/ftp/arxiv/papers/1311/1311.7388.pdf>
- [7]. Yacine Atif, “Building Trust in E-commerce”, 1089-7801/02 ©2002 IEEE.
- [8]. <http://www-users.cs.umn.edu/~desikan/publications/wmo.pdf>
- [9]. <http://www.ycjiaanguo.xyz/read-book/web-mining-applications-in-e-commerce-and-e-services.pdf>
- [10]. https://www.researchgate.net/publication/220520203_WEBKDD_2000_-_Web_Mining_for_E-Commerce
- [11]. Margaret H. Dunham, “Data Mining Introductory & Advanced Topics”, Pearson Education
- [12]. Ahmad Tasnim Siddiqui ,Arun Kumar Singh , “Secure E-business Transactions By Securing Web services” 978-0-7695-4853-1/12 © 2012 IEEE.
- [13]. Mrs. Pallavi Chandratre, Prof. Umesh Kulkarni “Implementation of trust ranking algorithm”, International Journal of Science Technology & Management, Volume No.04,, February 2015
- [14]. X. Fang and O. Sheng, Link Selector: A web mining approach to hyperlink selection for web portals, ACM Trans. Internet Tech. 4(2) (2004) 209–237
- [15]. B. Hay, G. Wets and K. Vanhoof, Mining navigation patterns using a sequence alignment method, Knowledge Inform. Syst. 6(2) (2004) 150–163